RESEARCH ARTICLE

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Oncogenic Potential is Related to Activating Effect of Cancer Single and Double Somatic Mutations in Receptor Tyrosine Kinases

Kosuke Hashimoto,[†] Igor B. Rogozin, and Anna R. Panchenko[‡]

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland

Communicated by Dominique Stoppa-Lyonnet

Received 23 January 2012; accepted revised manuscript 24 May 2012.

Published online 29 June 2012 in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.22145

ABSTRACT: Aberrant activation of receptor tyrosine kinases (RTKs) is a common feature of many cancer cells. It was previously suggested that the mechanisms of kinase activation in cancer might be linked to transitions between active and inactive states. Here, we estimate the effects of single and double cancer mutations on the stability of active and inactive states of the kinase domains from different RTKs. We show that singleton cancer mutations destabilize active and inactive states; however, inactive states are destabilized more than the active ones, leading to kinase activation. We show that there exists a relationship between the estimate of oncogenic potential of cancer mutation and kinase activation. Namely, more frequent mutations have a higher activating effect, which might allow us to predict the activating effect of the mutations from the mutation spectra. Independent evolutionary analysis of mutation spectra complements this observation and finds the same frequency threshold defining mutation hotspots. We analyze double mutations and report a positive epistasis and additional advantage of doublets with respect to cancer cell fitness. The activation mechanisms of double mutations differ from those of single mutations and double mutation spectrum is found to be dissimilar to the mutation spectrum of singletons.

Hum Mutat 33:1566–1575, 2012. Published 2012 Wiley Periodicals, Inc.*

KEY WORDS: cancer mutation; receptor tyrosine kinase; protein structure; kinase activation; mutation spectra; double mutations

Introduction

Receptor tyrosine kinases (RTKs) transduce signals from the extracellular matrix to the cytoplasm of a cell; they contain extracellular, transmembrane, and catalytic kinase domains and may include regulatory domains. RTK phosphorylation may lead to Ras activation and initiate the cascade of events, which lead to regulation of gene expression implicated in cell division, cellular homeostasis, and survival. It has been shown that kinases, especially RTKs, are frequently mutated in cancer; a large fraction of all alterations in cancer represent point mutations [Wood et al., 2007] and only a small fraction of mutations are inherited. Although somatic mutations contribute significantly to tumorigenesis, the large majority of them are considered to be neutral (so-called “passenger” mutations) and only a few of them are under positive selection in cancer cells (so-called “driver” mutations) [Greenman et al., 2007; Wood et al., 2007]. Certain cancer mutation hotspots have been detected in several RTKs, most of them located in the activation loop, P-loop, and Asp-Phe-Gly DFG loops. Moreover, driver mutations are more likely found to be associated with the functionally important regions in kinases than passenger mutations [Izarzugaza et al., 2009]. Although mutations in multiple genes have been observed in many cancers, interestingly, it was shown that properties of cancer mutations are more similar to Mendelian disease mutations than to complex disease mutations, pointing to the scenario when cancer progresses through a series of stepwise mutations each of which might provide some advantage to the tumor cells [Kaminker et al., 2007; Vogelstein and Kinzler, 1993]. Furthermore, according to the concept of “oncogene addiction,” cancer cells depend on the activity of a single or a few oncogenes for their proliferation and survival [Weinstein and Joe, 2008]. Finding oncogenes is not a trivial task and one approach includes finding significant driver mutations. Various statistical methods have been applied in the attempt to find positively selected mutants and distinguish driver from passenger mutations, but their predictive power remains limited and largely depends on the background mutation rate, which is difficult to determine for each sample [Torkamani et al., 2009].

The connection between cancer and kinase activation has been found fairly recently [Martin, 2004; Wan et al., 2004]. Such increased RTK activity may be caused by gene amplifications, enhanced transcription or translation, and also by mutations. Discovery of cancer-related mutations in several RTKs [Paez et al., 2004] and the analysis of their effect on kinase structure and activity revealed that some mutations may disturb autoinhibitory interactions and considerably accelerate catalysis [Dixit et al., 2009b; Yun et al., 2007]. Moreover, different mutations in RTKs were shown to result in different binding selectivity to cancer drugs. Structural studies of kinases in complexes with small molecule drugs indeed revealed different structural perturbations in response to cancer mutations, which might affect the equilibrium between active and inactive conformations [Engelman et al., 2007; Eswaran and Knapp, 2010; Greulich

---

Additional Supporting Information may be found in the online version of this article.

[†] Current address: Omics Science Center, RIKEN, Yokohama, Japan.

[‡] Correspondence to: Anna R. Panchenko, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Building 38A 8S814, Bethesda, MD 20894. E-mail: panch@ncbi.nlm.nih.gov

Contract grant sponsors: Intramural Research Program of the National Library of Medicine at the US National Institutes of Health; K.H. was supported by a JSPS Research Fellowship from the Japan Society for the Promotion of Science.

*Published 2012 Wiley Periodicals, Inc. * This article is a US Government work and, as such, is in the public domain of the United States of America.*

et al., 2005]. This reflects the importance of the analysis of structure and dynamics of kinases in understanding the mechanisms of cancer mutations.

The interconversion between active and inactive states in kinases is highly regulated and kinases differ in their mechanisms of activation and inactivation, specifically those processes that lead to conversion of inactive to active forms. Interestingly, it has been shown that despite commonly conserved features, inactive conformations might be more structurally diverse than active conformations and could be targeted selectively by small molecule inhibitors, among them cancer drugs [Jura et al., 2011; Schindler et al., 2000]. Activation mechanisms of kinases represent a spectrum with two extremes, with the *CDK1* (MIM# 116940) and *EGF* (MIM# 131550) receptors on one side that are activated by allosteric effectors, namely, by the formation of an asymmetric dimer in cases of Epidermal Growth Factor Receptor (EGFR) [Zhang et al., 2006] and by binding of cyclin in cases of CDK [De Bondt et al., 1993; Jeffrey et al., 1995]. At the other extreme, *SRC* (MIM# 190090) kinases are normally inhibited by binding of SH2 and SH3 domains and activated spontaneously when this interaction is disrupted [Moarefi et al., 1997]. In general, ligand binding to the extracellular region controls dimerization of kinase domains and subsequent cross-phosphorylation of tyrosine in the activation loop. Phosphorylated tyrosine can form an electrostatic contact with the basic residues and stabilize the active state of kinase, enabling phosphorylation of other tyrosine residues on the C-terminal tail, which in turn mediate binding of SH2 and PTB domains of downstream signaling molecules [Hubbard and Miller, 2007]. Moreover, it has been proposed that four hydrophobic residues forming the so-called "spine" contribute to the process of activation by coordinating the movements of the N- and C-lobes of the molecule [Kornev et al., 2006].

Although structural analyses showed significant differences in conformations between the active and inactive forms of kinase domains, only a few hotspot mutations have been analyzed by structural studies and dynamics simulations. The role of many rare as well as frequent mutations in the activation of kinases remains unclear. In this work, we examined the energetic effect of cancer mutations on both active and inactive conformations of kinase domain in relation to their oncogenic potential to quantify the coupling between cancer mutations, kinase stability, and activity. Oncogenic potential was measured as a number of samples where a given mutation was observed. Since cancer mutations can be observed not only as singletons but also in doublets or triplets, we analyzed mutation patterns for single and double mutations in the kinase domain (the juxtamembrane, JM, region was also included if its structure was available). In accord with other studies, we showed that cancer mutations had an activating effect on RTKs. Singleton cancer mutations overall destabilized both active and inactive states; however, inactive states were destabilized more than the active ones so that active states were more populated. Interestingly, more frequently observed mutations had a higher activating effect for both single and double mutations, which might allow one to predict the activating effect of the mutations from their mutation spectra. Independent evolutionary analysis of mutation spectra complemented this observation and found the same frequency threshold defining the mutation hotspots. Moreover, for many double mutations, we found a positive epistasis or nonadditive effect, which pointed to the additional advantage of doublets for the tumor cell compared with singletons. The evolutionary analysis also demonstrated that nonsynonymous single and multiple mutations in RTKs occurred more often than expected by chance and led to selective advantages for tumor cells. In addition, the mutation spectrum of multiple mutations was found to be different from the spectra of singletons, which hints at different underlying mechanisms of their origin.

Materials and Methods

Examining the Frequency Distribution of Cancer Mutations

We derived all mutations available for 58 different human RTKs from the COSMIC database v49 Release [Forbes et al., 2008], which stores somatic mutations of cancer cells extracted from the primary literature. Amino acid sequences of the 58 RTKs and mutation numbering were obtained from the COSMIC database (Supp. Table S1). We divided the sequences into five regions: the extracellular region, transmembrane domain, JM domain, kinase domain, and C-terminal tail. First, we determined the boundaries of the transmembrane domain. Second, to determine the boundaries of the kinase domain, we generated a multiple sequence alignment of 58 sequences using the MAFFT program [Katoh and Toh, 2008] and mapped secondary structure elements onto the alignment using a crystal structure of the kinase domain of EGFR. We defined the first residue of the first β-strand of the N-lobe as the first site of the kinase domain and the last residue of the last α-helix of the C-lobe as the terminal site of the domain. Finally, the extracellular region was defined between the N-terminal end and the transmembrane domain, the JM region—between the transmembrane domain and the kinase domain and the C-terminal tail—between the end of the kinase domain and the end of the protein. For each region, we counted the number of unique mutations and mutation sites, and calculated the number of unique mutation sites per residue in the protein. The evolutionary tree of 58 RTKs was constructed based on the multiple alignment of the kinase domain with the neighbor-joining method [Saitou and Nei, 1987] using MEGA5.

Assessing the Effect of Mutations on RTKs in Active and Inactive States

We obtained all crystal structures for RTKs available in the manually curated KEGG pathway database [Kanehisa et al., 2004], and manually compiled a list of structures of the kinase domain in active and inactive states (altogether 18 structures). We then calculated the change in unfolding free energy ($\Delta\Delta G$) of single amino acid substitutions on these structures. Perturbations in $\Delta\Delta G$ caused by mutations may result in changes of equilibrium constant and concentrations of RTKs in different conformations. $\Delta\Delta G$ was calculated using two modules of FoldX version 3.0 [Guerois et al., 2002], which is among the top available methods to estimate the effect of mutations on protein stability; it reaches 0.64 sensitivity and 0.43 specificity [Khan and Vihinen, 2010] of prediction and reports a correlation coefficient between experimental and computed $\Delta\Delta G$ values in the range of 0.5–0.8 [Guerois et al., 2002; Potapov et al., 2009; Zhang et al., 2012].

First, we identified strained torsion angles and Van der Waals' clashes in the original structure and optimized side chains to provide a repaired structure (RepairPDB module). Then, using the BuildModel module, we optimized the configurations of the neighboring side chains of the mutation site and calculated the difference in stability ($\Delta\Delta G$) between the repaired native structure and the mutant structure.

For each mutation and corresponding amino acid substitution, we calculated a $\Delta\Delta\Delta G$ value, which represents the difference in effect of the mutation on the stability of the active and inactive states:

$$\Delta\Delta\Delta G = \Delta\Delta G\,(\text{inactive}) - \Delta\Delta G\,(\text{active}) \tag{1}$$

$$\Delta\Delta G = \Delta G(\text{wt}) - \Delta G(\text{mut}) \tag{2}$$

where $\Delta G(\text{wt})$ and $\Delta G(\text{mut})$ were the free energy of unfolding for wild type and structures with mutations, respectively. The unfolded state of wild type and mutants were considered to be similar as described in previous studies [Zhang et al., 2010, 2011]. The $\Delta\Delta\Delta G$ value was calculated only when the mutation could be mapped on both structures in the active and inactive states. We also extracted combinations of two or more concurrent mutations from our dataset, which we call thereafter "multiple" mutations. The $\Delta\Delta G$ values of the observed multiple mutations were calculated using the BuildModel module of FoldX. Negative and positive $\Delta\Delta G$ values corresponded to stabilizing and destabilizing effects of mutations, respectively. In addition, we sampled the background $\Delta\Delta G$ distribution by selecting all possible single amino acid substitutions caused by single nucleotide substitutions for all sites in a protein (for analysis of single mutations) and by randomly selecting 1,000 pairs of amino acid substitutions (for analysis of double mutations).

We also defined the positive epistatic (superadditive) effect of multiple mutations as follows:

$$SA = \Delta\Delta\Delta G\,(AB) - \Delta\Delta\Delta G(A) - \Delta\Delta\Delta G\,(B) \tag{3}$$

Here, $\Delta\Delta\Delta G(AB)$, $\Delta\Delta\Delta G(A)$, and $\Delta\Delta\Delta G(B)$ describe the effect of double and single mutations on activation, respectively. Positive SA values correspond to the positive epistasis, namely, when the shift toward the active state upon introducing both mutations is higher than the sum of the effects of the single mutations. All data for single and double mutations produced in this study is available at ftp://ftp.ncbi.nih.gov/pub/panch/RTK/.

Predicting Mutation Hotspots

Mutation hotspot prediction in this study was based on a threshold (Sh) value for the number of mutations in a mutable site. The threshold and resulting hotspot sites were calculated for each mutation spectrum (mutation spectrum represents a distribution of a number of position with a given mutation frequency) separately using classification analysis described previously [Glazko et al., 1998; Rogozin et al., 2001]. For this purpose, the CLUSTERP program was used (ftp://ftp.bionet.nsc.ru/pub/biology/dbms/CLUSTERM.ZIP), which is based on the simulation, expectation, and maximization subclass approach. The algorithm tries to classify the mutation sites according to different mutation probabilities, and each site should belong to only one class. The mutation spectrum of each class is approximated by the Poisson distribution and an overall mutational spectrum is regarded as a mixture of Poisson distributions. Variations in mutation frequencies among sites of the same class are assumed to be due to random reasons (since mutation probability is the same for all sites in one class), but differences between mutation frequencies among sites from different classes are statistically significant. A class with the highest mutation frequency is called a hotspot class. The process which separates the mutation spectrum into classes is iterative and each iteration includes simulation, maximization, and estimation procedures.

Analyzing Doublet Mutations

To study the properties of multiple mutation spectra, we used a sampling procedure repeated 1,000 times. Each generated set of pseudo-multiple mutations had the size equal to the set of multiple mutations and mutations were randomly taken from the set of single mutations. The resulting set of generated pseudo-multiple mutations was compared with the observed set of multiple mutations. A Monte Carlo modification of the Pearson χ^2 test of spectra homogeneity [Adams and Skopek, 1987] was used to compare distributions of multiple and pseudo-multiple mutations along the protein sequences. Calculations were done using the program COLLAPSE [Khromov-Borisov et al., 1999]. We used the above sampling procedure to estimate the expected number of doublets and the significance of their overrepresentation and underrepresentation. The number of doublets averaged over 1,000 trials was used as an expected value. Statistical significance of overrepresented doublets was estimated using *P* value, which referred to the probability to find the observed or larger number of a given doublet purely by chance ($P[S \geq O]$). If *P* value was less than 0.05, this pair was considered overrepresented. The same logic was used for the analysis of underrepresented doublets where the *P* value $P(O \geq S)$ was calculated. This procedure was applied to pairs of mutations with non-zero expected values only. This severely limits the number of analyzed pairs, however increases the reliability of the analysis.

We also employed an analogy to the Hardy–Weinberg model [Crow, 1999; Hardy, 1908]. The Hardy–Weinberg principle is frequently used for allele and genotype frequencies: zygotic genotype frequencies are predictable from gamete frequencies, assuming random mating. In the case of double mutations, we assume that frequencies of nonsynonymous and synonymous substitutions are allele frequencies and those substitutions are randomly combined in doublets. We found that the observed number of "doublet heterozygosity" (doublets that contain one nonsynonymous and one synonymous substitution) is significantly lower than expected under Hardy–Weinberg equilibrium (Supp. Table S2).

We used BioRuby [Goto et al., 2010] and the Entrez Programming Utilities [Sayers et al., 2011] to facilitate data manipulation and analyses throughout this study and Cytoscape [Shannon et al., 2003] for the network visualization.

Results

The Kinase and JM Domains have the Largest Density of Cancer Mutations

We obtained 9,607 nonsynonymous mutations observed in different cancer samples from the COSMIC database [Forbes et al., 2008], including 1,060 unique mutations from 841 unique mutation sites. Out of these 9,607 mutations, zygosity information was available for 1,534 mutations and 93% of them were heterozygous, namely contained the mutation in only one allele of the locus. The mutations were distributed over a wide range of human RTKs (Supp. Fig. S1), and *EGFR* and *KIT* (MIM# 164920), well known cancer causative genes, had more than 100 unique mutation sites. The majority of mutations and mutation sites (73%) were observed only in one cancer sample (patient), whereas about 5% of them were observed in 10 or more samples (Supp. Fig. S2). The most frequently observed mutation in the dataset was the p.Leu858Arg mutation of *EGFR* (observed in 2,299 samples). These repeated mutations are more likely to be driver mutations in contrast to the mutations observed

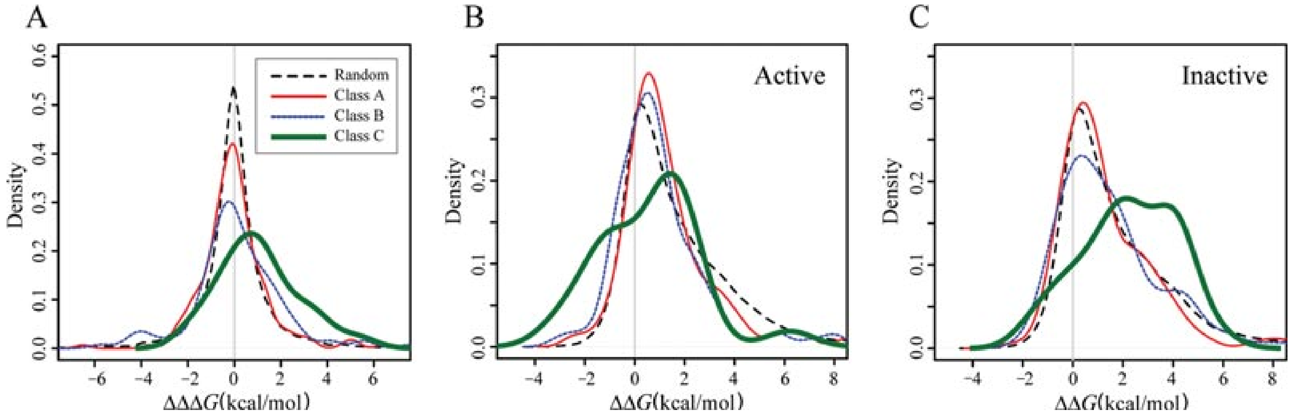

Figure 1. Effect of single mutations on stability of the kinase domain. **A:** The distribution of $\Delta\Delta\Delta G$ for unique mutations of class A is shown in red/thin line (187 mutations), class B in blue/stippled line (110 mutations), class C in green/bold line (21 mutations), and all possible random mutations caused by single nucleotide substitutions are shown in black dashed line. **B:** The distribution of $\Delta\Delta G$ of active states. **C:** The distribution of $\Delta\Delta G$ of inactive states. The probability density functions were smoothed using the Gaussian kernel.

only in one sample, the large fraction of which might correspond to passenger mutations.

All mutations were classified depending on their oncogenic potential (which was estimated here as a frequency of samples where they were observed) into class "A"—for those observed in one sample, class "B"—for those observed in two to nine samples, and class "C"—for 10 or more samples. This classification is consistent with our evolutionary analysis described below. We observed that the kinase domain and JM region had the largest density of mutations out of all RTK regions, next came the extracellular domain, whereas mutations in the membrane domain and the C-terminal tail were relatively rare (Supp. Fig. S3). This result is consistent with the previous studies [Dixit et al., 2009b; Izarzugaza et al., 2009] although mutations were classified differently in the present study. Moreover, from Supp. Fig S3, one can see that the oncogenic potential varies greatly depending on the region of RTKs. For example, the percentage of mutation sites increased from 42% for class A to 57–59% for class B/C sites for the kinase domain and almost doubles (from 9% to 15–20%) in the case of the JM domain. In contrast, the number of mutation sites decreased with the oncogenic potential for the extracellular domain, which points to many rare mutations in this region.

The Effect of Single Mutations on Stability and Activity of Kinase Domain

Since the majority of mutations in class B and C are located in the kinase domain and their frequencies per residue are also high, we measured the effect of the mutations on the stability and activity of the kinase domain (in some cases, the JM region was also included). We collected six pairs of structures of the same RTK kinase in the active and inactive states and structures from six additional RTK kinases available either in the active or inactive conformations (altogether seven crystal structures in the active state and 11 structures in the inactive state, Supp. Table S1). In total, 344 unique mutations were mapped onto active structures and 419 mutations were mapped onto inactive structures, of which 318 were mapped onto both of them.

Figure 1A shows the distribution of $\Delta\Delta\Delta G$ values, which represents the difference in the unfolding free energies between the active and inactive states upon introducing the amino acid substitution corresponding to a cancer mutation (see *Materials and Methods*). Positive $\Delta\Delta\Delta G$ values correspond to the tendency for the activation by mutations when the equilibrium between active and inactive states is shifted toward active states. As can be seen from Figures 1A and 2, the mean values of these distributions increase from −0.01 kcal/mol for class A, to +0.15 kcal/mol for class B, and +1.06 kcal/mol for class C cancer mutations. The distribution of class C is significantly shifted to the positive side ($P = 0.001$, one-sided Wilcoxon rank-sum test), compared with the background distribution (see *Materials and Methods*). Cancer mutations from the A and B classes do not show a significant effect on the stability of active or inactive states compared with random mutations ($P = 0.71$ and 0.25, respectively). This indicates that the frequently observed mutations have a different effect (the mean value of $\Delta\Delta\Delta G$ would be close to zero if the energetic effects of mutations on the active and inactive conformations were comparable) on the active and inactive states, leading to the shift of equilibrium from inactive to active states.

We analyzed separately the energetic effect of cancer mutations on the active and inactive conformations. Figure 1B and C shows distributions of $\Delta\Delta G$ values for the active and inactive states, respectively. All three distributions for active states are shifted to the positive values, indicating that cancer mutations destabilize active structures, however destabilize them significantly less than the random mutations ($P = 0.033$ for A, 0.0006 for B, and 0.017 for C classes, respectively). As to the inactive structures, unlike cancer mutations from the A and B classes, only mutations from class C destabilize inactive structures significantly compared with random mutations ($P = 0.025$) (Fig. 1C). In short, mutations with high oncogenic potential have a tendency to destabilize inactive states to a greater degree than active states, which may cause aberrant activations of RTKs in cancer cells (Fig. 2). This effect is more pronounced for the *KIT* family of RTKs.

Assuming that homozygosity or heterozygosity of RTK mutations should not affect the activity since mutant oncogene alleles are typically dominant, we indeed did not find any difference in $\Delta\Delta\Delta G$ distributions between heterozygous and homozygous mutations even though the most frequent mutations such as p.Leu858Arg in *EGFR* and p.Asp816Val or p.Leu576Pro in *KIT* were sometimes observed as homozygous.

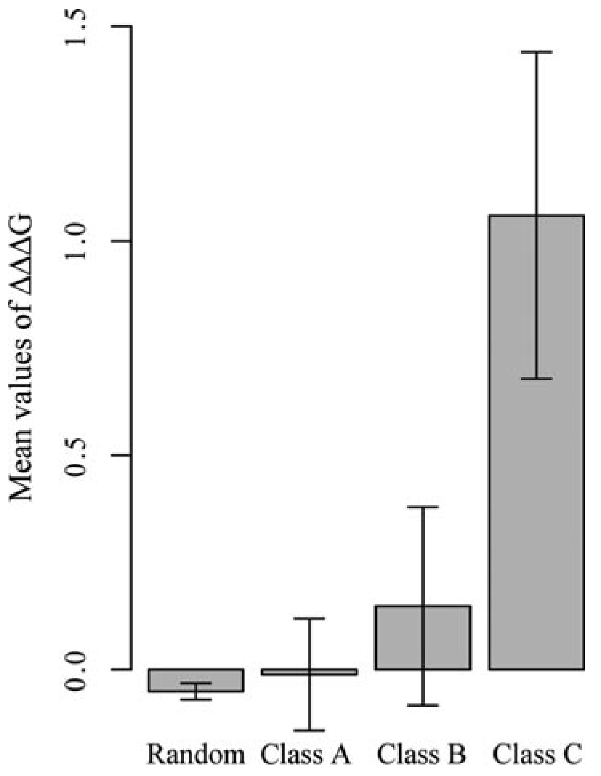

Figure 2. The relationship between oncogenic and activation potentials. Mean values and standard errors of $\Delta\Delta\Delta G$ plotted for each class of unique mutations.

Analysis of Single Mutation Spectra

In this work, we analyzed two mutation spectra of the *EGFR* and *KIT* genes. The numbers of single mutations in *EGFR* and *KIT* were 2,680 and 2,034, respectively. Analysis of the mutation spectrum of the *EGFR* gene using CLUSTERP revealed four classes of sites (see *Materials and Methods*). The first class includes obvious "cold" sites with mutation frequencies (number of samples the same mutation was observed) less than or equal to 4, the second class includes sites with the mutation frequency less than 10, the third class from 10 to 26, and the fourth class comprises obvious hotspot sites (mutation frequency > 49). The distribution of observed and expected mutation frequencies is shown in Supp. Figure S4A. The second class does not contain hotspot sites since numerous sites with no mutations or just one mutation were also included in this class, while several obvious hotspots were present in the third class of sites. Thus, 10 mutations were chosen as the threshold value (Sh) for determining the mutation hotspot sites. Similar results were obtained for the *KIT* gene with 2,034 mutations (Supp. Fig. S4B); analysis of four predicted classes of sites suggested the same threshold of 10 mutations, which is consistent with our class C derived in the previous section based on the analysis of oncogenic and activation potentials.

The great majority of *EGFR* and *KIT* mutations (>95%) occurred in nonsynonymous codon sites and, accordingly, resulted in amino acid replacements in *EGFR* and *KIT* genes. The excess of nonsynonymous over synonymous substitutions was statistically significant ($P < 10^{-70}$ by the Fisher exact test) when compared with the random expectation under the neutral evolution model. Namely, using the modified Nei–Gojobori method, we estimated that approximately 73% of the substitutions were expected to occur in nonsynonymous sites in the neutral regime. Preponderance of amino acid replacement over silent substitutions is the signature of positive (directional) selection [Hurst, 2002]. The large excess of nonsynonymous substitutions seen in *EGFR* and *KIT* is similar to *TP53* (MIM# 191170) and some other cancer-related genes [Glazko et al., 2006] and is an unequivocal indication that, in tumors, the *EGFR* and *KIT* genes evolve under positive selection and preferentially accumulate mutations that lead to selective advantages for the tumor cells.

The Effect of Multiple Mutations on Stability and Activity of Kinase Domain

We found 210 unique multiple mutations from 17 different RTKs, and about 73% of them were from the *EGFR* or *KIT* families (Supp. Fig. S5). The most frequent multiple mutation was the combination of p.Thr790Met and p.Leu858Arg in the *EGFR* protein, which was observed in 48 samples. We mapped 97 multiple mutations that consisted of 92 double mutations and five triple mutations on structures in active and inactive states (so-called "doublets"). Figure 3A shows distributions of $\Delta\Delta\Delta G$ values for multiple mutations and a background distribution of a randomly chosen 1,000 pairs of mutations. The distribution corresponding to cancer mutations is significantly shifted to positive values compared with random doublet mutations ($P \ll 0.01$, one-sided Wilcoxon rank-sum test) pointing to a tendency for activation for doublets. Interestingly, the distribution of $\Delta\Delta G$ for active kinase states is significantly shifted to negative values ($P \ll 0.01$), compared with the background distribution, implying the stabilization of the active state by double cancer mutations compared with random mutations (Fig. 3B), while the inactive state does not seem to be significantly destabilized by doublets ($P = 0.33$) (Fig. 3C).

Moreover, we found a positive epistasis for double mutations (Fig. 4). Indeed, the effect of multiple mutations on shifting the population of kinase conformations toward active ones is higher than a total of individual mutations ($P \ll 0.01$, one-sided Wilcoxon rank-sum test), suggesting that, overall, double mutations have a synergistic effect. This trend is especially pronounced for double mutations observed in more than one sample (Fischer exact test $P = 0.02$). In addition, we do not observe any correlation between $\Delta\Delta\Delta G$ values and the spatial distances between two mutations in the structures (Supp. Fig. S6), indicating that synergistic multiple mutations are not necessarily close to each other in space providing for their direct interactions.

To understand the mechanism of multiple mutations, we focused on doublets found specifically in *EGFR* and *KIT*. We constructed a mutation network, where nodes represented mutation sites and edges corresponded to double mutations. The network shows that L858 of *EGFR* is clearly the biggest hub, which connects to more than 20 nodes/mutation sites in doublets (Fig. 5), although this hub is even smaller than expected based on the single mutation frequencies (see *Analysis of Doublet Mutation Spectra*). *EGFR* also has some hotspots that connect to more than five nodes, such as G719, S768, and L861, and all of them are in the kinase domain. On the contrary, *KIT* has many more concurrent mutations between residues in the kinase domain and the JM domain, showing many cliques and connections between multiple sites. This might reflect the difference between their regulation mechanisms. At the same time, there are seven sites aligned in both *KIT* and *EGFR* that are involved in double mutations (shown in magenta in Fig. 5).

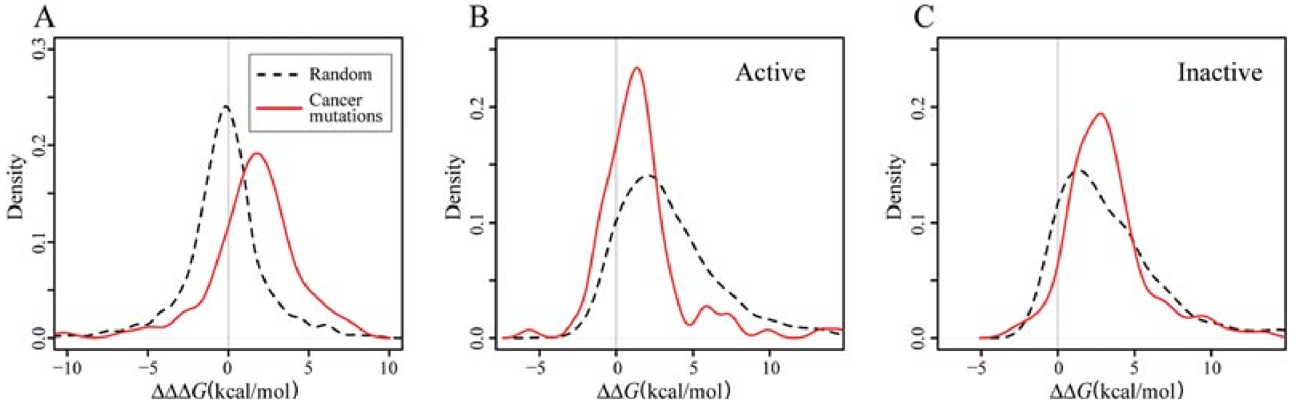

Figure 3. Effect of multiple mutations on stability of the kinase domain. **A**: The distribution of $\Delta\Delta\Delta G$ for observed multiple mutations in red/solid line and random mutations in black dashed line. **B**: The distribution of $\Delta\Delta G$ of active states. **C**: The distribution of $\Delta\Delta G$ of inactive states. The probability density functions were smoothed using the Gaussian kernel.

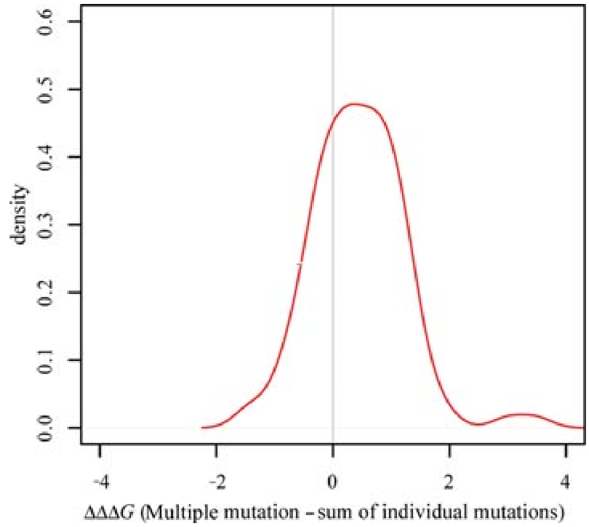

Figure 4. Positive epistasis of double mutations. Distribution of SA (difference between $\Delta\Delta\Delta G$ [multiple mutations]) and $\Delta\Delta\Delta G$ (sum of corresponding mutations).

Analysis of Doublet Mutation Spectra

The numbers of doublet mutations in *EGFR* and *KIT* were 207 and 76 pairs, respectively. Similar to the single mutations, the great majority of multiple mutations (>88%) occurred in nonsynonymous codon sites implying that doublets are under positive selection. To compare the properties of multiple mutation spectra to single mutations, we used a sampling procedure (see *Materials and Methods*). The observed and simulated spectra of double mutations (generated from the spectra of single mutations) were significantly different ($P \ll 0.01$ for *EGFR* and *KIT*), which points to the different mechanisms underlying single and double mutation spectra. The major difference between observed and simulated mutation spectra for the *EGFR* gene came from the significantly smaller number of multiple mutations in the major hotspot position 858 (270 observed mutations versus 373 expected mutations, $P \ll 0.01$ by the Fisher exact test). Several positions with elevated frequency of multiple mutations, for example, the positions 709 (15 mutations), 719 (42 mutations), and 768 (22 mutations) in the *EGFR* gene were also observed. We listed doublets with a significant overrepresentation and underrepresentation in Table 1.

We also compared the frequency of double mutations in *EGFR* and *KIT* to the *TP53* gene [Meng et al., 1999] and found that the frequency of nonsynonymous doublets was significantly higher in the *EGFR* and *KIT* genes compared with *TP53* (Supp. Table S2). This result strongly suggests that multiple mutations in *EGFR* and *KIT* are driven by positive selection. Multiple mutations in *TP53* are likely to be the result of transient mutagenesis [Drake et al., 2005] and are unlikely to be under positive selection [Rodin and Rodin, 1998] and thus can be used as a null model [Drake et al., 2005; Rodin and Rodin, 1998]. Additional analysis using the Hardy–Weinberg principle (see *Materials and Methods*) also suggests that strong positive selection is an important driving force for RTK doublets. Under this model, we compared the observed level of "doublet heterozygosity" (doublets that contain one nonsynonymous and one synonymous substitution) to what we expect under Hardy–Weinberg equilibrium. The observed "doublet heterozygosity" ($2pq = 51$) is significantly lower than expected ($2pq = 82$) (Supp. Table S2); this discrepancy is usually attributed to preferential fixation of double nonsynonymous mutations due to positive selection for *EGFR* and *KIT*.

Interestingly, quite a few RTK mutations and mutation sites in our set are found only as a part of doublet mutations, namely, 49% of mutations and 30% of mutation sites are observed only in doublets but not in singletons. For example, the p.Glu709Gly mutations in *EGFR*, observed in seven different patients, always appear with another mutation rather than as a single mutation. These mutations are likely to be secondary mutations.

Discussion

It has been previously shown that cancer mutations may activate RTKs, sometimes in a ligand-independent way, and the mechanisms of kinase activation in cancer might be linked to transitions between the active and inactive states [Yun et al., 2007; Zhang et al., 2006]. The crystal structures of the *EGFR* p.Leu858Arg mutant showed, for example, that this mutation prevents the activation loop from adopting the inactive conformation [Yun et al., 2007]. It was suggested that the effect of the secondary *EGFR* p.Thr790Met mutation facilitates interconversion between the inactive and active conformations and enhances the stability of the active conformation relative

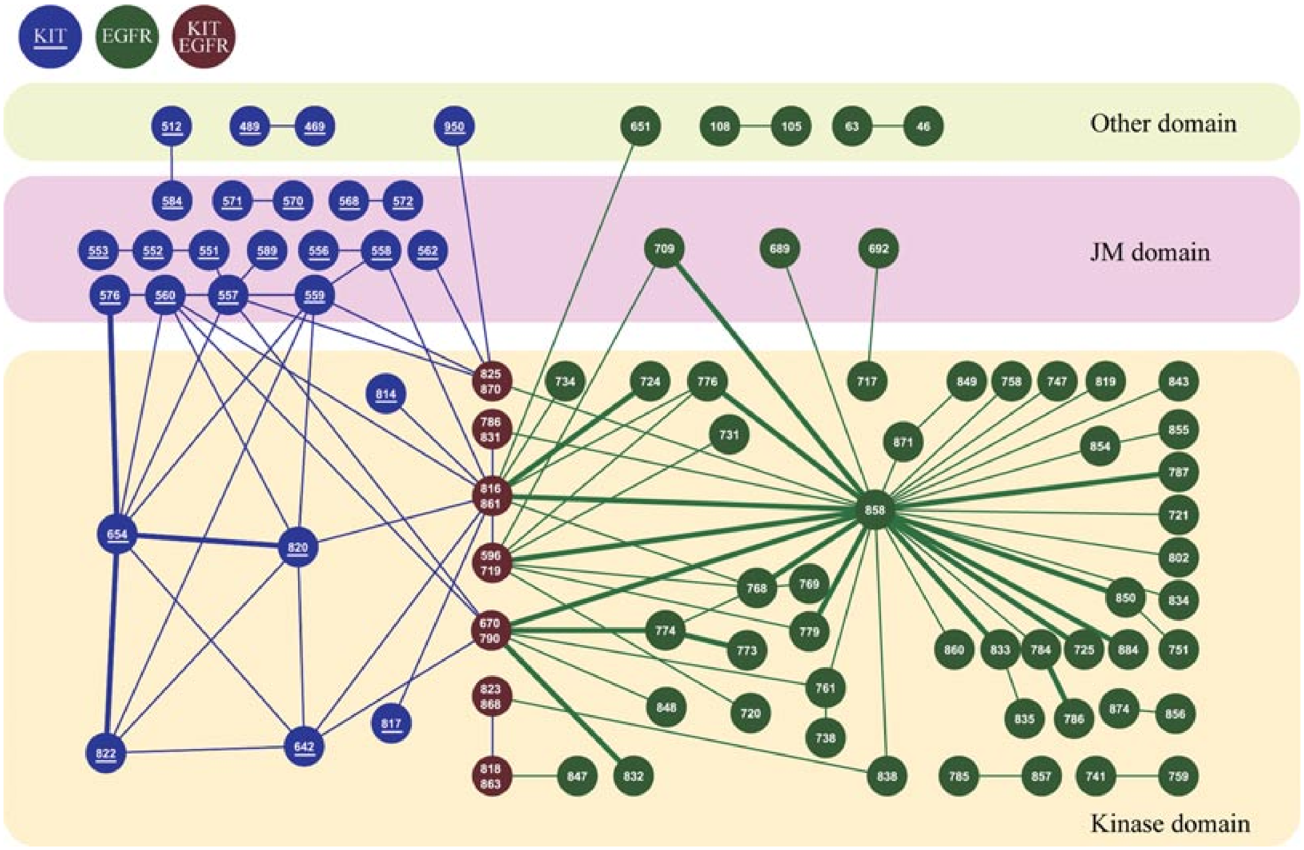
KIT
EGFR
KIT EGFR
Other domain
JM domain
Kinase domain
512
489
469
950
651
108
105
63
46
584
571
570
568
572
553
552
551
589
556
558
562
576
560
557
559
709
689
692
825 870
814
786 831
734
724
776
731
717
871
849
758
747
819
843
854
855
787
654
816 861
858
721
820
596 719
802
768
769
850
834
670 790
774
773
779
860
833
784
725
884
751
823 868
848
720
761
835
786
874
856
738
817
822
642
818 863
847
832
838
785
857
741
759

Figure 5. The mutation network of EGFR and KIT for different RTK regions. Each node represents one mutation site where at least one mutation can occur. The mutation sites of *KIT* are shown in blue/gray, *EGFR* in green/black, and aligned sites are shown in maroon/white. Residue numbers are given inside the nodes. The edge connects two sites that are concurrently mutated. Those edges with positive epistasis are shown in bold.

Table 1. Significantly Overrepresented and Underrepresented Double Mutations in EGFR and KIT

Protein	Mutation 1	Mutation 2	Observed	Expected	*P* value
	Overrepresented pairs				
EGFR	p.Leu833Val	p.Leu858Arg	11	0.38	0.001
EGFR	p.Gly719Cys	p.Ser768Ile	6	0.17	0.005
EGFR	p.Glu709Gly	p.Leu858Arg	6	0.90	0.009
EGFR	p.Thr790Met	p.Leu858Arg	47	28.74	0.022
KIT	p.Leu576Pro	p.Val654Ala	4	0.18	0.031
	Underrepresented pairs				
EGFR	p.Gly719Ser	p.Leu858Arg	3	21.75	0.004
EGFR	p.Leu858Arg	p.Leu861Val	1	16.1	0.002
KIT	p.Val560Asp	p.Asp816Val	1	6.6	0.024

All mutation numbering was obtained from the COSMIC database. All mutations were mapped to a single version of each gene sequence, and are available in the Download section of the COSMIC database.

to the inactive one [Yun et al., 2008]. More recently, the modeling of autoinhibited conformations and the effect of several frequent hotspot mutations revealed enhanced mobility near mutation sites, which disrupted the local stabilizing interactions and in some cases allosterically altered the distribution of locally frustrated sites and destabilized the inactive form [Dixit et al., 2009a, 2009b; Dixit and Verkhivker, 2011]. It was shown that *EGFR* p.Thr790Met and *EGFR* p.Leu858Arg mutations also lead to the enhanced stability of the active state [Dixit and Verkhivker, 2009].

In this study, we analyzed both the active and inactive states of the kinase domain from structures of different RTKs and estimated the effect of single and multiple cancer mutations on their stability. In accordance with the previous studies, using much larger dataset of mutations, we showed that, overall, single cancer mutations destabilized inactive states. We also showed that single cancer mutations destabilized active states of RTK, but to a lesser degree than the random mutations; moreover, to a lesser degree than the single cancer mutations destabilized inactive states. This led to kinase activation. The exception was the *EGFR* p.Thr790Met mutation, which did not have a considerable overall effect ($\Delta\Delta\Delta G$ = –0.5 kcal/mol) alone and exhibited an activating effect as shown later as a double mutation together with *EGFR* p.Leu858Arg (see next section). Although destabilizing effect on both active and inactive states might potentially lead to misfolding and aggregation, the changes in $\Delta\Delta G$ of about 1–3 kcal/mol might not compromise an overall fitness of a protein [Tokuriki and Tawfik, 2009]. The homozygosity state of the mutation did not affect the obtained results, consistent with the premises that mutant oncogene alleles were typically dominant.

Moreover, we tried to answer the question if one can predict the activating effect of single or multiple mutations from their mutation spectra. Importantly, we showed that there exists a relationship between the statistics-based estimate of oncogenic potential of mutation and its activation effect calculated based on thermodynamics principles. Namely, more frequent mutations have a somewhat higher activating effect. This effect is not linear, although, and for frequent mutations from more than 10 samples, the activity increases radically upon introducing mutations. From the Boltzmann distribution, one can estimate the fraction of proteins with a given stability or stability effect of mutations ($\Delta\Delta G$), which would follow a simple logistic function behavior. Indeed, previously, the sigmoidal relationship was revealed between the destabilizing effects of mutations, leading to monogenic diseases and the severity of the diseases [Yue et al., 2005]. In addition, recently an association was shown between the frequency of mutations and their potential functional impact calculated based on their evolutionary conservation [Reva et al., 2011]. Our analysis of thermodynamic properties is nicely complemented by the evolutionary analysis, which uncovered four populations of mutations with different underlying mutation rates and we showed that mutation hotspots (potential driver mutations) can be reliably defined from the samples with frequencies higher than 10. It should be mentioned that although we analyzed local changes produced by cancer mutations in several RTK families differing by regulatory mechanisms, we observed the common activating effect of cancer mutations in relation to its oncogenic potential. Thus, our analysis complements previous observations. It implies that although increased RTK activity may be caused by gene amplifications, enhanced transcription, or translation, both single and double point mutations might play a key role in kinase activation in cancer.

Moreover, a significant fraction of cancer-associated mutations comes in doublets or triplets. The origin of multiple mutations is still not very well understood, they can originate either through defects in DNA replication/repair systems or arise locally through transient mechanisms [Chen et al., 2009, 2012; Matsuda et al., 2001; Pavlov et al., 2006; Seidman et al., 1987; Stone et al., 2009]. We found that 7.7% and 3.7% of all single cancer mutations of *EGFR* and *KIT*, respectively, represent double cancer mutations, compared with the previously reported 6% of double mutations in *EGFR* in lung cancers [Chen et al., 2008]. These numbers are substantially larger than the fraction of multiple somatic mutations in the lacI transgene in mouse somatic tissues (~1%) [Hill et al., 2004]. Furthermore, the fraction of multiple mutations in neighboring positions (tandem mutations) of the *lacI* gene was found to be higher compared with multiple mutations separated by one or more nucleotides [Hill et al., 2004], whereas the opposite tendency was found in the *EGFR* gene where tandem mutations were rare (results not shown). In general, an excess of tandem mutations is a signature of various error-prone DNA polymerases and is expected to have a distinct DNA context specificity [Matsuda et al., 2001; Pavlov et al., 2006; Stone et al., 2009]. This is consistent with the distinct context properties of tandem lacI mutations in mouse somatic tissues and a lack of significant differences for spectra of single and nontandem multiple lacI mutations [Buettner et al., 2000; Hill et al., 2004]. However, the mutation spectrum of multiple mutations in our study was found to be different from the spectrum of singletons, which hints at different underlying mechanisms of their origin and suggests a role of clonal selection for multiple substitutions [Bazykin et al., 2004]. Although we did not find any evidence for prevalence of double substitutions to contact each other in three-dimensional protein structures of kinase monomers, such contacts may occur between different subunits of homodimers since dimerization is crucial for RTK functions in a cell. Further analysis is needed to study the effect of mutations on stability and function of protein complexes since many disease-related mutations may disrupt protein interactions [Schuster-Bockler and Bateman, 2008; Teng et al., 2009].

Similar to single mutations, double mutations were shown to activate the kinase domain. Moreover, for many doublets, a positive epistasis has been found, which points to the additional advantage of doublets for the tumor cell compared with singletons. Some doublets have a drug resistance effect, one of the classical examples is the p.Thr790Met+p.Leu858Arg double mutant of *EGFR* showing strong resistance to gefitinib [Tam et al., 2009]. According to our study, this double mutation is predicted to have a potential activating effect ($\Delta\Delta\Delta G$ = +1.9 kcal/mol) and quite high oncogenic potential (observed in 47 samples). Moreover, it exhibits a positive epistasis effect (SA = +1.14 kcal/mol). Some other mutations with high oncogenic potential and positive epistasis include p.Leu833Val + p.Leu858Arg ($\Delta\Delta\Delta G$ = +3.85 kcal/mol; SA = +1.4 kcal/mol) and p.Leu858Arg + p.Glu884Lys ($\Delta\Delta\Delta G$ = +4.57 kcal/mol and SA = +1.07 kcal/mol) of *EGFR*. The latter E884K mutation disrupts an ion pair between K884 and R958 and may change the downstream signaling significantly although its connection to the p.Leu858Arg mutation remains unclear [Tang et al., 2009]. In addition, we observed many doublets, which were not found as single mutations (about half of all doublets). One might hypothesize that such mutations are secondary mutations causing the previously mentioned differences between the single and double mutation spectra.

Analyses of mutation spectra and activating effects of different cancer mutations have important prognostic implications. Namely, it has been shown that some RTK activating mutants have been associated with a better survival prognosis and better response to RTK inhibitors [Jackman et al., 2009]. Indeed, cancer cells exhibiting mutant kinases become critically dependent on certain pathways (so-called "oncogene addiction" [Weinstein, 2002]). For example, mutant *EGFR*s selectively activates Akt and signal transduction/activator of transcription signaling pathways, which in turn promote cell survival. It explains the effectiveness of gefitinib, inhibiting critical antiapoptotic pathways in lung cancers with mutant *EGFR* genotypes [Sordella et al., 2004]. In our study, we attempted to link the stability of RTKs with their oncogenic potential and differential activity, which combined with other data on phosphorylation patterns for each mutant may provide insight into the mechanisms of activation of different pathways by cancer mutations and may help to design effective cancer drugs.

Acknowledgments

We thank Jeyanthy Eswaran, Joshua Cherry, and Tom Madej for very helpful discussions.

Disclosure Statement: The authors declare no conflict of interest.

References

Adams WT, Skopek TR. 1987. Statistical test for the comparison of samples from mutational spectra. J Mol Biol 194:391–396.

Bazykin GA, Kondrashov FA, Ogurtsov AY, Sunyaev S, Kondrashov AS. 2004. Positive selection at sites of multiple amino acid replacements since rat-mouse divergence. Nature 429:558–562.

Buettner VL, Hill KA, Scaringe WA, Sommer SS. 2000. Evidence that proximal multiple mutations in Big Blue transgenic mice are dependent events. Mutat Res 452:219–229.

Chen JM, Ferec C, Cooper DN. 2009. Closely spaced multiple mutations as potential signatures of transient hypermutability in human genes. Hum Mutat 30:1435–1448.

Chen JM, Ferec C, Cooper DN. 2012. Transient hypermutability, chromothripsis and replication-based mechanisms in the generation of concurrent clustered mutations. Mutat Res 750:52–59.

Chen Z, Feng J, Buzin CH, Sommer SS. 2008. Epidemiology of doublet/multiplet mutations in lung cancers: evidence that a subset arises by chronocoordinate events. PLoS One 3:e3714.

Crow JF. 1999. Hardy, Weinberg and language impediments. Genetics 152:821–825.

De Bondt HL, Rosenblatt J, Jancarik J, Jones HD, Morgan DO, Kim SH. 1993. Crystal structure of cyclin-dependent kinase 2. Nature 363:595–602.

Dixit A, Torkamani A, Schork NJ, Verkhivker G. 2009a. Computational modeling of structurally conserved cancer mutations in the RET and MET kinases: the impact on protein structure, dynamics, and stability. Biophys J 96:858–874.

Dixit A, Verkhivker GM. 2009. Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations. PLoS Comput Biol 5:e1000487.

Dixit A, Verkhivker GM. 2011. The energy landscape analysis of cancer mutations in protein kinases. PLoS One 6:e26071.

Dixit A, Yi L, Gowthaman R, Torkamani A, Schork NJ, Verkhivker GM. 2009b. Sequence and structure signatures of cancer mutation hotspots in protein kinases. PLoS One 4:e7485.

Drake JW, Bebenek A, Kissling GE, Peddada S. 2005. Clusters of mutations from transient hypermutability. Proc Natl Acad Sci U S A 102:12849–12854.

Engelman JA, Zejnullahu K, Gale CM, Lifshits E, Gonzales AJ, Shimamura T, Zhao F, Vincent PW, Naumov GN, Bradner JE, Althaus IW, Gandhi L, et al. 2007. PF00299804, an irreversible pan-ERBB inhibitor, is effective in lung cancer models with EGFR and ERBB2 mutations that are resistant to gefitinib. Cancer Res 67:11924–11932.

Eswaran J, Knapp S. 2010. Insights into protein kinase regulation and inhibition by large scale structural comparison. Biochim Biophys Acta 1804:429–432.

Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. 2008. The catalogue of somatic mutations in cancer (COSMIC). Curr Protoc Hum Genet Chapter 10:unit10.11.

Glazko GB, Milanesi L, Rogozin IB. 1998. The subclass approach for mutational spectrum analysis: application of the SEM algorithm. J Theor Biol 192:475–487.

Glazko GV, Babenko VN, Koonin EV, Rogozin IB. 2006. Mutational hotspots in the TP53 gene and, possibly, other tumor suppressors evolve by positive selection. Biol Direct 1:4.

Goto N, Prins P, Nakao M, Bonnal R, Aerts J, Katayama T. 2010. BioRuby: bioinformatics software for the Ruby programming language. Bioinformatics 26:2617–2619.

Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, et al. 2007. Patterns of somatic mutation in human cancer genomes. Nature 446:153–158.

Greulich H, Chen TH, Feng W, Janne PA, Alvarez JV, Zappaterra M, Bulmer SE, Frank DA, Hahn WC, Sellers WR, Meyerson M. 2005. Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. PLoS Med 2:e313.

Guerois R, Nielsen JE, Serrano L. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320:369–387.

Hardy GH. 1908. Mendelian proportions in a mixed population. Science 28:49–50.

Hill KA, Wang J, Farwell KD, Scaringe WA, Sommer SS. 2004. Spontaneous multiple mutations show both proximal spacing consistent with chronocoordinate events and alterations with p53-deficiency. Mutat Res 554(:223–240.

Hubbard SR, Miller WT. 2007. Receptor tyrosine kinases: mechanisms of activation and signaling. Curr Opin Cell Biol 19:117–123.

Hurst LD. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 18:486.

Izarzugaza JM, Redfern OC, Orengo CA, Valencia A. 2009. Cancer-associated mutations are preferentially distributed in protein kinase functional sites. Proteins 77:892–903.

Jackman DM, Miller VA, Cioffredi LA, Yeap BY, Janne PA, Riely GJ, Ruiz MG, Giaccone G, Sequist LV, Johnson BE. 2009. Impact of epidermal growth factor receptor and KRAS mutations on clinical outcomes in previously untreated non-small cell lung cancer patients: results of an online tumor registry of clinical trials. Clin Cancer Res 15:5267–5273.

Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich NP. 1995. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature 376:313–320.

Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T, Kuriyan J. 2011. Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. Mol Cell 42:9–22.

Kaminker JS, Zhang Y, Waugh A, Haverty PM, Peters B, Sebisanovic D, Stinson J, Forrest WF, Bazan JF, Seshagiri S, Zhang Z. 2007. Distinguishing cancer-associated missense mutations from common polymorphisms. Cancer Res 67:465–473.

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res 32:D277–D280.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9:286–298.

Khan S, Vihinen M. 2010. Performance of protein stability predictors. Hum Mutat 31:675–684.

Khromov-Borisov NN, Rogozin IB, Pegas Henriques JA, de Serres FJ. 1999. Similarity pattern analysis in mutational distributions. Mutat Res 430:55–74.

Kornev AP, Haste NM, Taylor SS, Eyck LF. 2006. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. Proc Natl Acad Sci U S A 103:17783–17788.

Martin GS. 2004. The road to Src. Oncogene 23:7910–7917.

Matsuda T, Bebenek K, Masutani C, Rogozin IB, Hanaoka F, Kunkel TA. 2001. Error rate and specificity of human and murine DNA polymerase eta. J Mol Biol 312:335–346.

Meng L, Lin L, Zhang H, Nassiri M, Morales AR, Nadji M. 1999. Multiple mutations of the p53 gene in human mammary carcinoma. Mutat Res 435:263–269.

Moarefi I, LaFevre-Bernt M, Sicheri F, Huse M, Lee CH, Kuriyan J, Miller WT. 1997. Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. Nature 385:650–653.

Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, et al. 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304:1497–1500.

Pavlov YI, Shcherbakova PV, Rogozin IB. 2006. Roles of DNA polymerases in replication, repair, and recombination in eukaryotes. Int Rev Cytol 255:41–132.

Potapov V, Cohen M, Schreiber G. 2009. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein Eng Des Sel 22:553–560.

Reva B, Antipin Y, Sander C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39:e118.

Rodin SN, Rodin AS. 1998. Strand asymmetry of CpG transitions as indicator of G1 phase-dependent origin of multiple tumorigenic p53 mutations in stem cells. Proc Natl Acad Sci U S A 95:11927–11932.

Rogozin I, Kondrashov F, Glazko G. 2001. Use of mutation spectra analysis software. Hum Mutat 17:83–102.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, Feolo M, Fingerman IM, et al. 2011. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 39:D38–D51.

Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. 2000. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science 289:1938–1942.

Schuster-Bockler B, Bateman A. 2008. Protein interactions in human genetic diseases. Genome Biol 9:R9.

Seidman MM, Bredberg A, Seetharam S, Kraemer KH. 1987. Multiple point mutations in a shuttle vector propagated in human cells: evidence for an error-prone DNA polymerase activity. Proc Natl Acad Sci U S A 84:4944–4948.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504.

Sordella R, Bell DW, Haber DA, Settleman J. 2004. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. Science 305:1163–1167.

Stone JE, Kissling GE, Lujan SA, Rogozin IB, Stith CM, Burgers PM, Kunkel TA. 2009. Low-fidelity DNA synthesis by the L979F mutator derivative of *Saccharomyces cerevisiae* DNA polymerase zeta. Nucleic Acids Res 37:3774–3787.

Tam IY, Leung EL, Tin VP, Chua DT, Sihoe AD, Cheng LC, Chung LP, Wong MP. 2009. Double EGFR mutants containing rare EGFR mutant types show reduced in vitro response to gefitinib compared with common activating missense mutations. Mol Cancer Ther 8:2142–2151.

Tang Z, Jiang S, Du R, Petri ET, El-Telbany A, Chan PS, Kijima T, Dietrich S, Matsui K, Kobayashi M, Sasada S, Okamoto N, et al. 2009. Disruption of the EGFR E884-R958 ion pair conserved in the human kinome differentially alters signaling and inhibitor sensitivity. Oncogene 28:518–533.

Teng S, Madej T, Panchenko A, Alexov E. 2009. Modeling effects of human single nucleotide polymorphisms on protein–protein interactions. Biophys J 96:2178–2188.

Tokuriki N, Tawfik DS. 2009. Stability effects of mutations and protein evolvability. Curr Opin Struct Biol 19:596–604.

Torkamani A, Verkhivker G, Schork NJ. 2009. Cancer driver mutations in protein kinase genes. Cancer Lett 281:117–127.

Vogelstein B, Kinzler KW. 1993. The multistep nature of cancer. Trends Genet 9:138–141.

Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D, Marais R; Cancer Genome Project. 2004. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell 116:855–867.

Weinstein IB. 2002. Cancer. Addiction to oncogenes–the Achilles heal of cancer. Science 297:63–64.

Weinstein IB, Joe A. 2008. Oncogene addiction. Cancer Res 68:3077–3080.

Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, et al.. 2007. The genomic landscapes of human breast and colorectal cancers. Science 318:1108–1113.

Yue P, Li Z, Moult J. 2005. Loss of protein structure stability as a major causative factor in monogenic disease. J Mol Biol 353:459–473.

Yun CH, Boggon TJ, Li Y, Woo MS, Greulich H, Meyerson M, Eck MJ. 2007. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. Cancer Cell 11:217–227.

Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, Meyerson M, Eck MJ. 2008. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci U S A 105:2070–2075.

Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J. 2006. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell 125:1137–1149.

Zhang Z, Norris J, Schwartz C, Alexov E. 2011. In silico and in vitro investigations of the mutability of disease-causing missense mutation sites in spermine synthase. PLoS One 6:e20373.

Zhang Z, Teng S, Wang L, Schwartz CE, Alexov E. 2010. Computational analysis of missense mutations causing Snyder–Robinson syndrome. Hum Mutat 31:1043–1049.

Zhang Z, Wang L, Gao Y, Zhang J, Zhenirovskyy M, Alexov E. 2012. Predicting folding free energy changes upon single point mutations. Bioinformatics 28:664–671.